10-11-2006

Essential Elements of a Defense-Review of DNA Testing Results

Dan E. Krane
Wright State University - Main Campus, dan.krane@wright.edu

Essential elements of a defense-review of DNA testing results

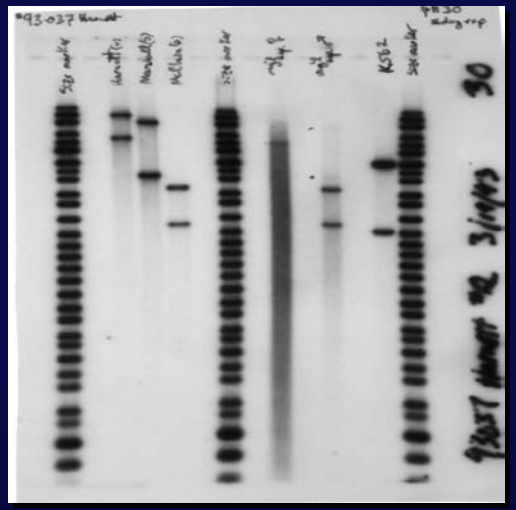
Dan E. Krane, Wright State University, Dayton, OH

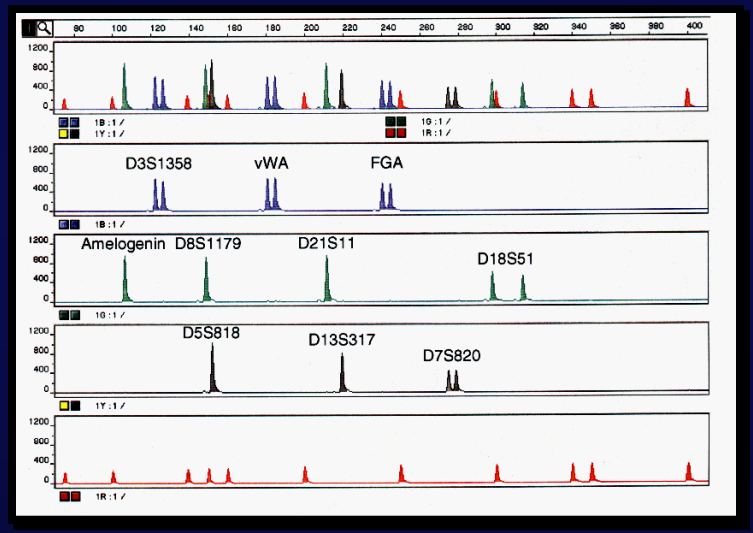
Forensic Bioinformatics
(www.bioforensics.com)

The science of DNA profiling is sound.

But, not all of DNA profiling is science.

Three generations of DNA testing





RFLP
AUTORAD
Allele = BAND

DQ-alpha
TEST STRIP
Allele = BLUE DOT

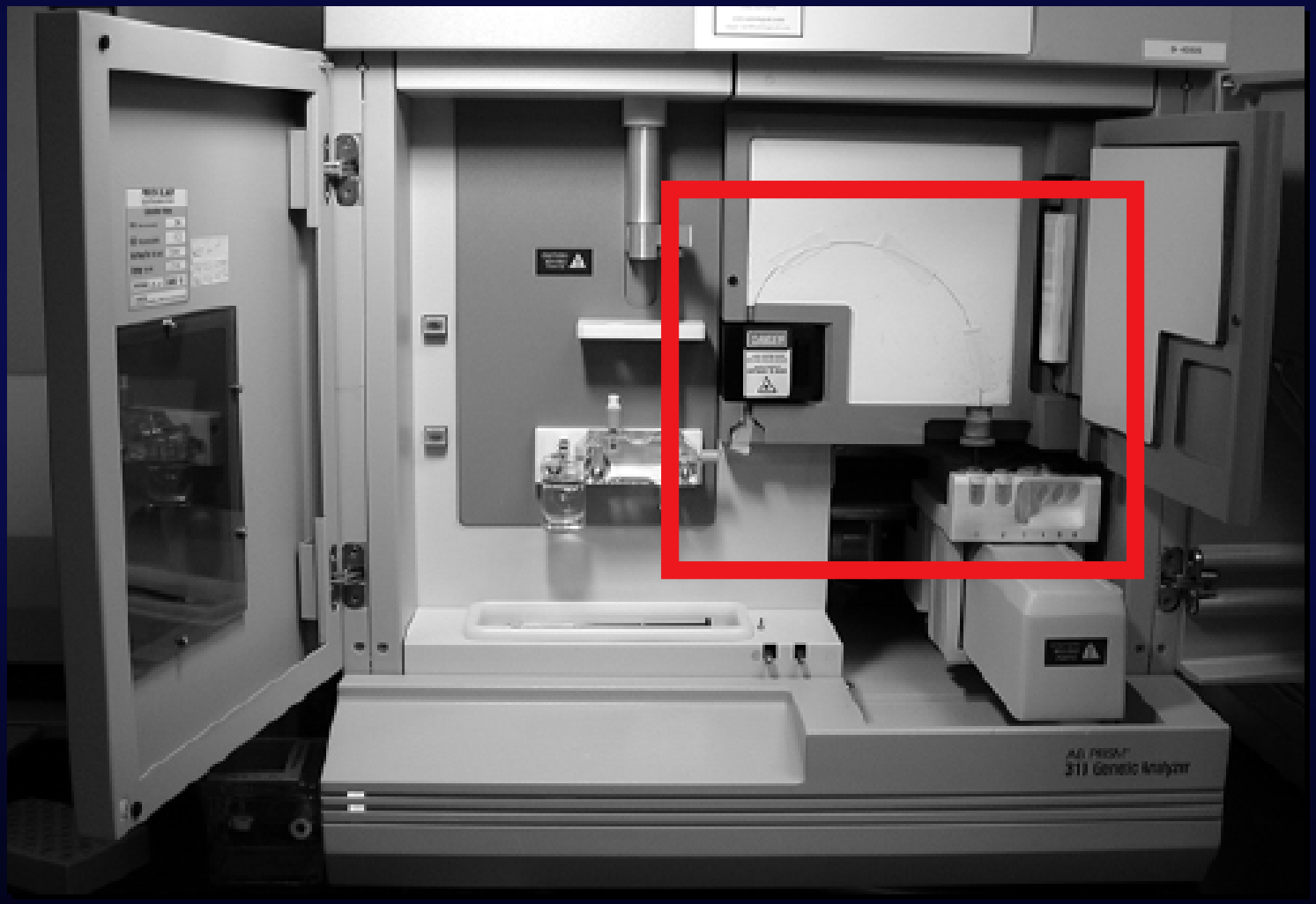
Automated STR
ELECTROPHEROGRAM
Allele = PEAK

DNA content of biological samples:

Type of sample	Amount of DNA
Blood	30,000 ng/mL
stain 1 cm^2 in area	200 ng
stain 1 mm^2 in area	2 ng
Semen	250,000 ng/mL
Postcoital vaginal swab	0 - 3,000 ng
Hair	
plucked	1 - 750 ng/hair
shed	1 - 12 ng/hair
Saliva	5,000 ng/mL
Urine	1 - 20 ng/mL

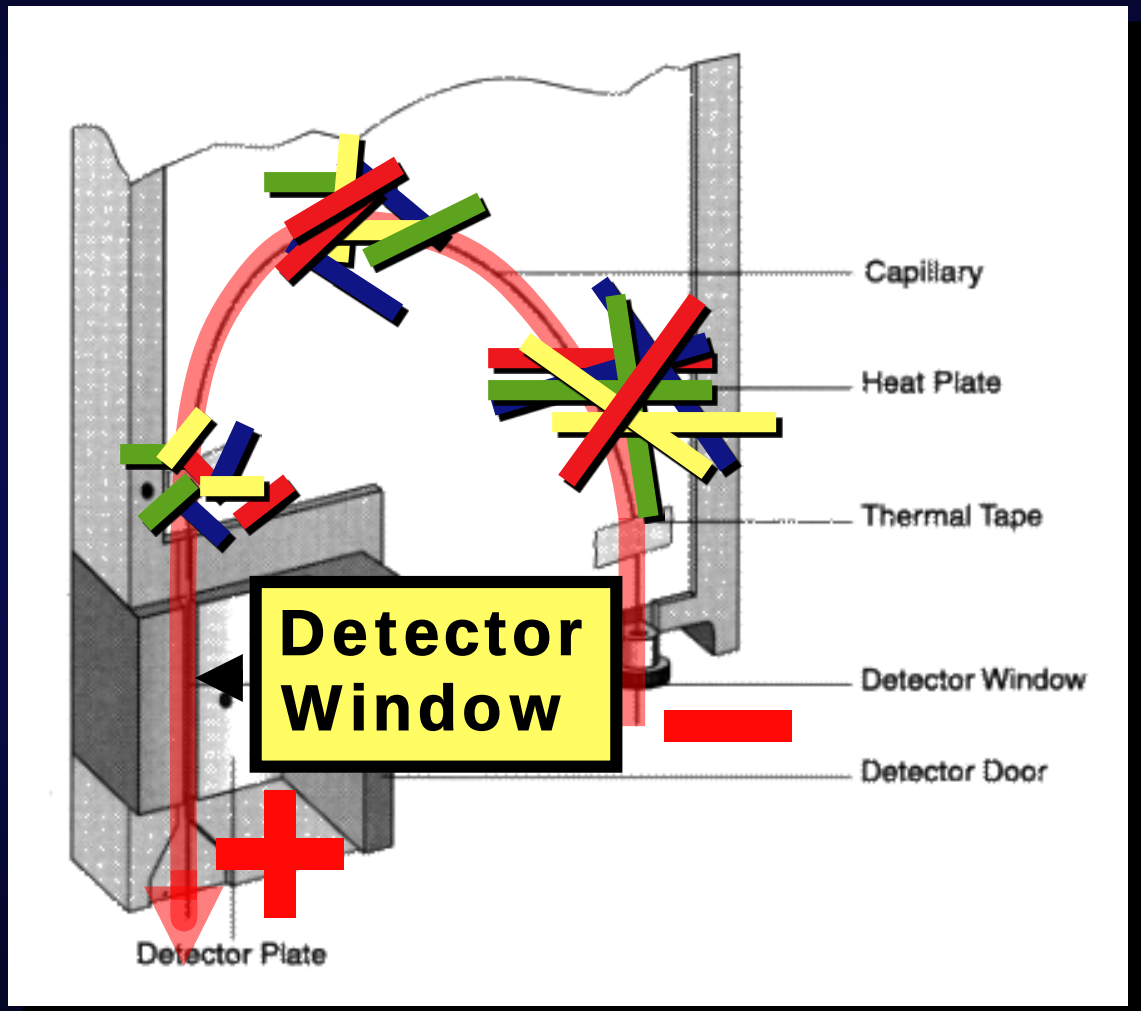
Automated STR Test

The ABI 310 Genetic Analyzer

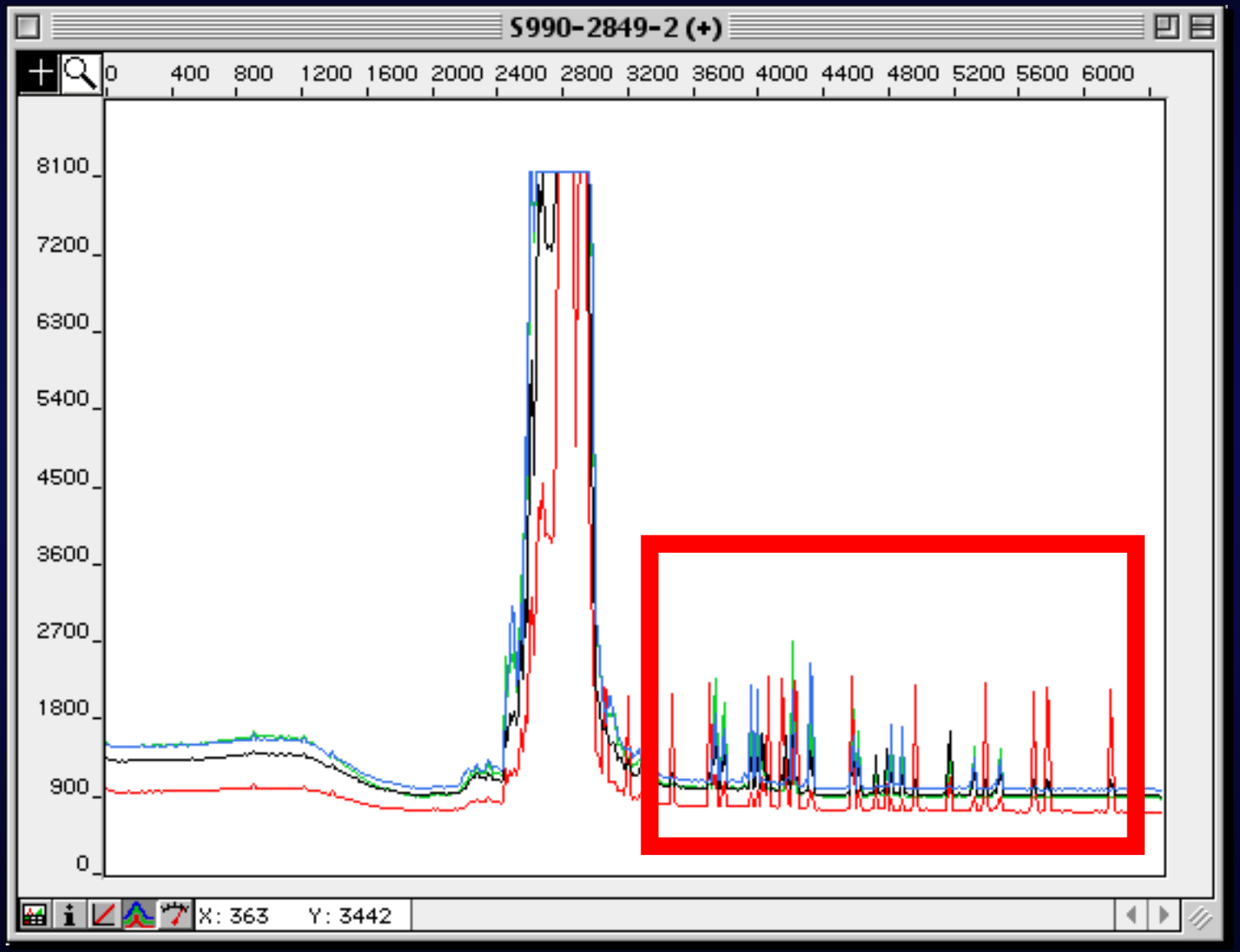


ABI 310 Genetic Analyzer: Capillary Electrophoresis

- Amplified STR DNA injected onto column
- Electric current applied
- DNA pulled towards the positive electrode
- DNA separated out by size:
  - Large STRs travel slower
  - Small STRs travel faster
- Color of STR detected and recorded as it passes the detector



Profiler Plus: Raw data



Statistical estimates: the product rule

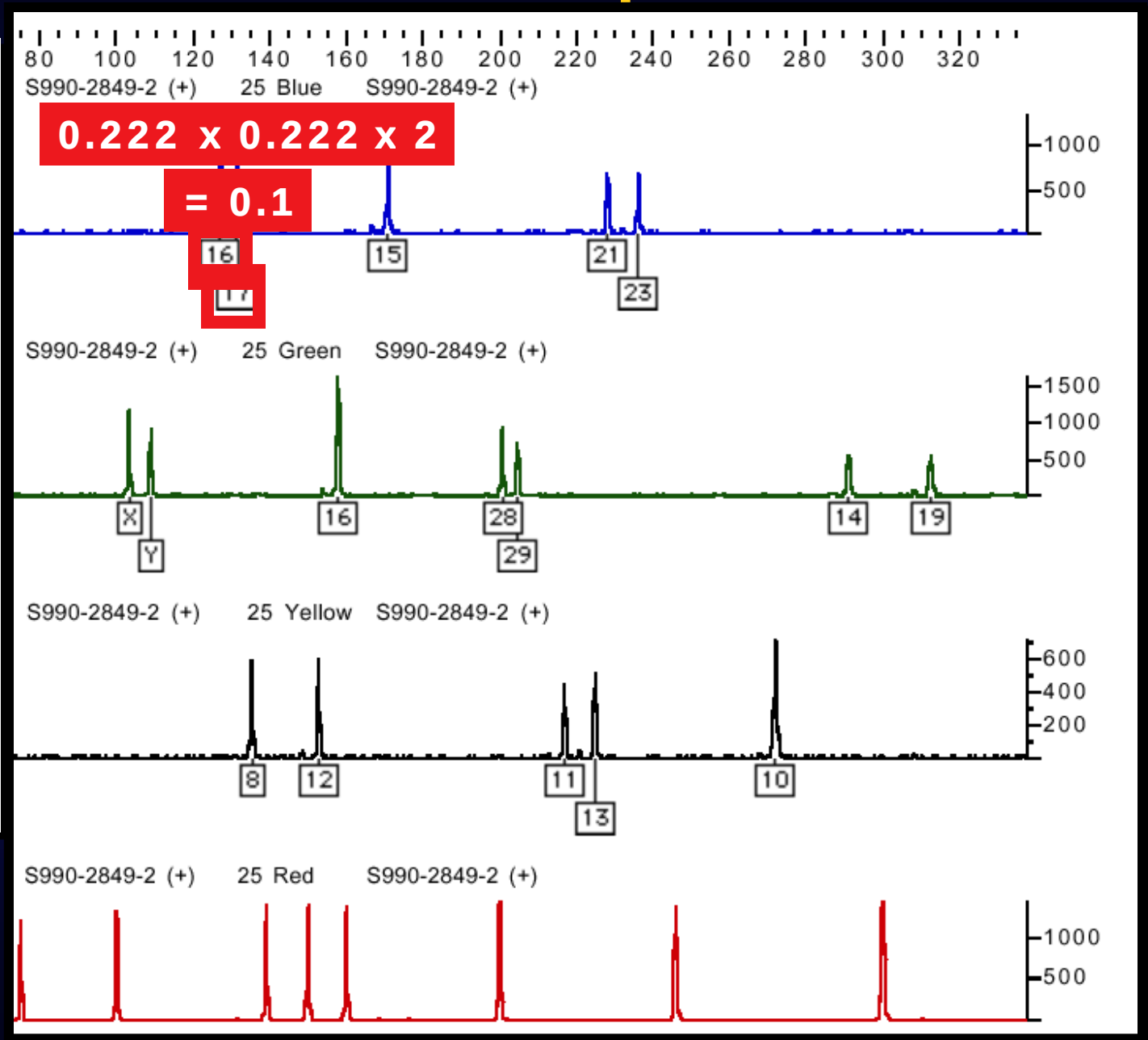
Allele Frequencies

Locus D3S1358
Race Caucasian
(N = 203)

Allele	Frequency
12	0.012
13	0.012
14	0.140
[redacted]	[redacted]
19	0.012

Locus vWA
Race Caucasian
(N = 196)

Allele	Frequency
11	0.012
12	0.012
13	0.012
14	0.102
15	0.082



Statistical estimates: the product rule

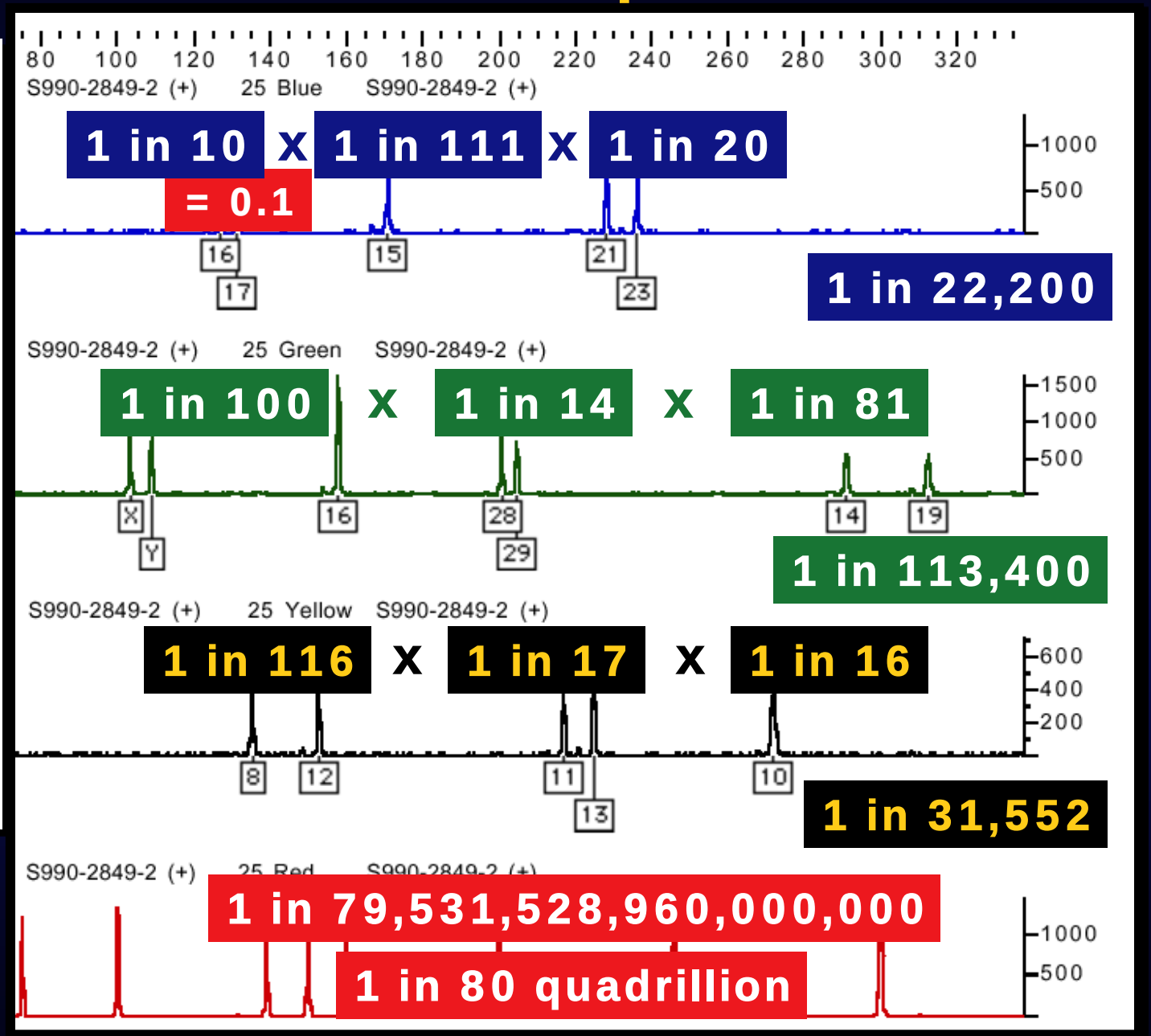
Allele Frequencies

Locus D3S1358
Race Caucasian
(N = 203)

Allele	Frequency
12	0.012
13	0.012
14	0.140
15	0.246
16	0.222
17	0.222
18	0.163
19	0.012

Locus vWA
Race Caucasian
(N = 196)

Allele	Frequency
11	0.012
12	0.012
13	0.012
14	0.102
15	0.082



What more is there to say after you have said: “The chance of a coincidental match is one in 80 quadrillion?”

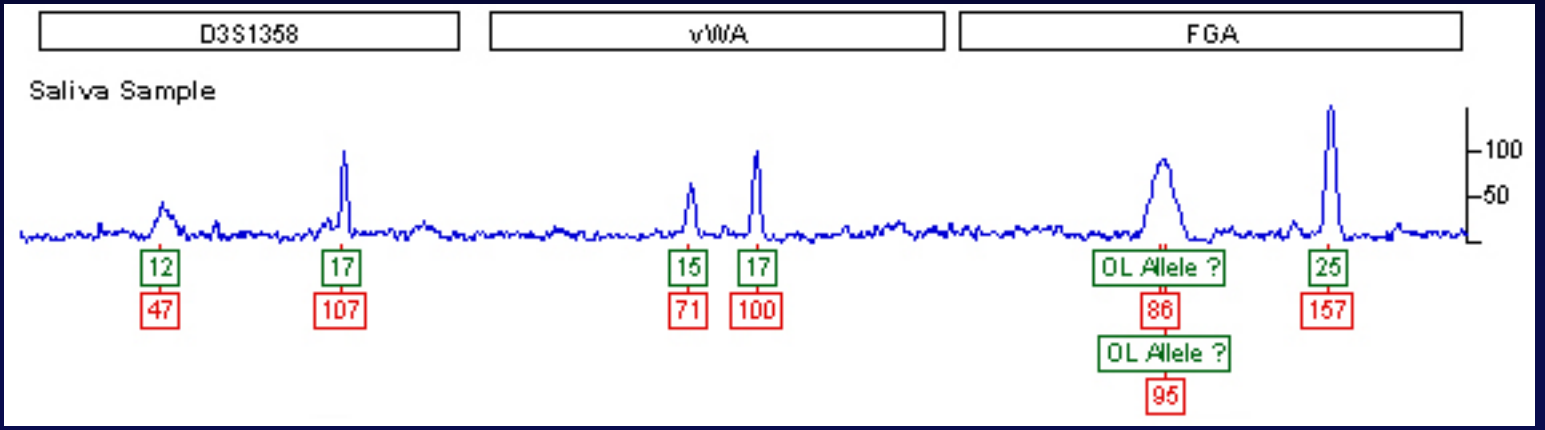
What more is there to say after you have said: “The chance of a coincidental match is one in 80 quadrillion?”

- Two samples really do have the same source
- Samples match coincidentally
- An error has occurred

The science of DNA profiling is sound.

But, not all of DNA profiling is science.

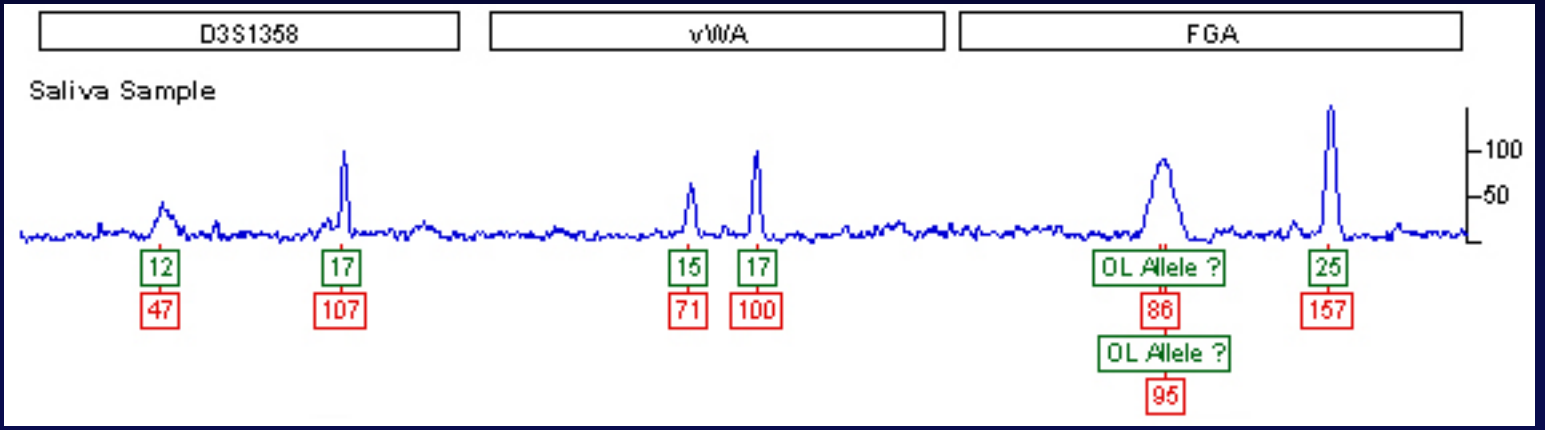
Opportunities for subjective interpretation?



Can “Tom” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25

Opportunities for subjective interpretation?

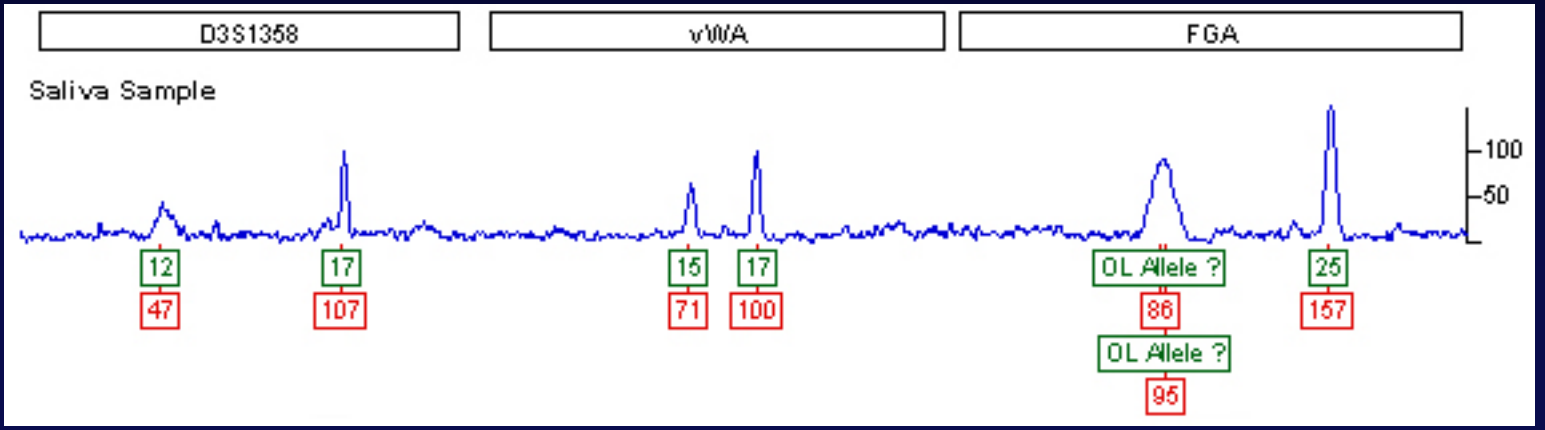


Can “Tom” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25

No -- the additional alleles at D3 and FGA are “technical artifacts.”

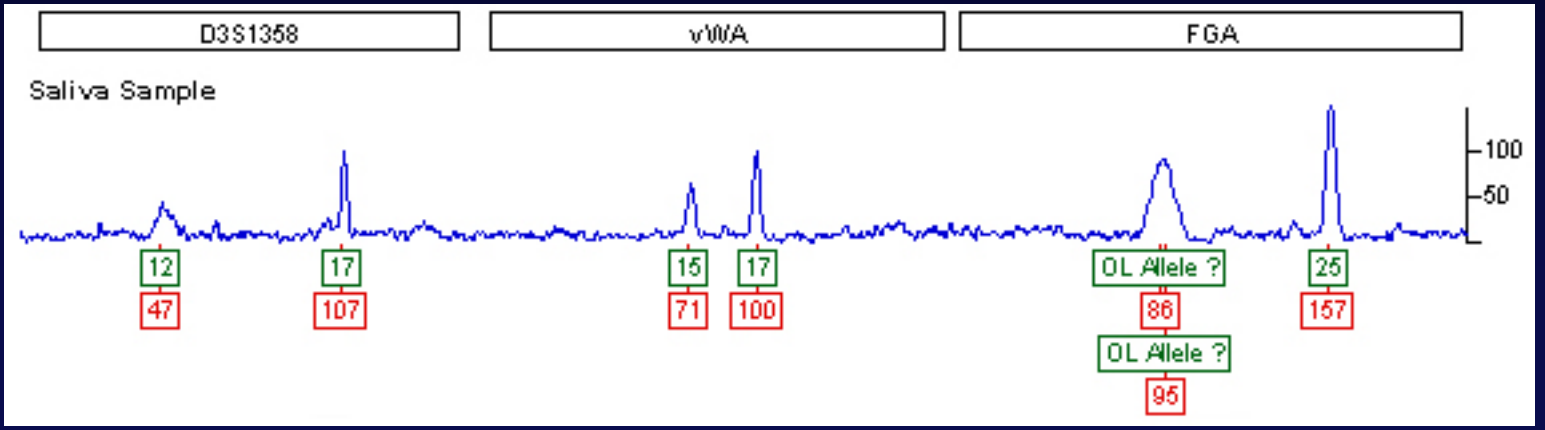
Opportunities for subjective interpretation?



Can “Dick” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25

Opportunities for subjective interpretation?

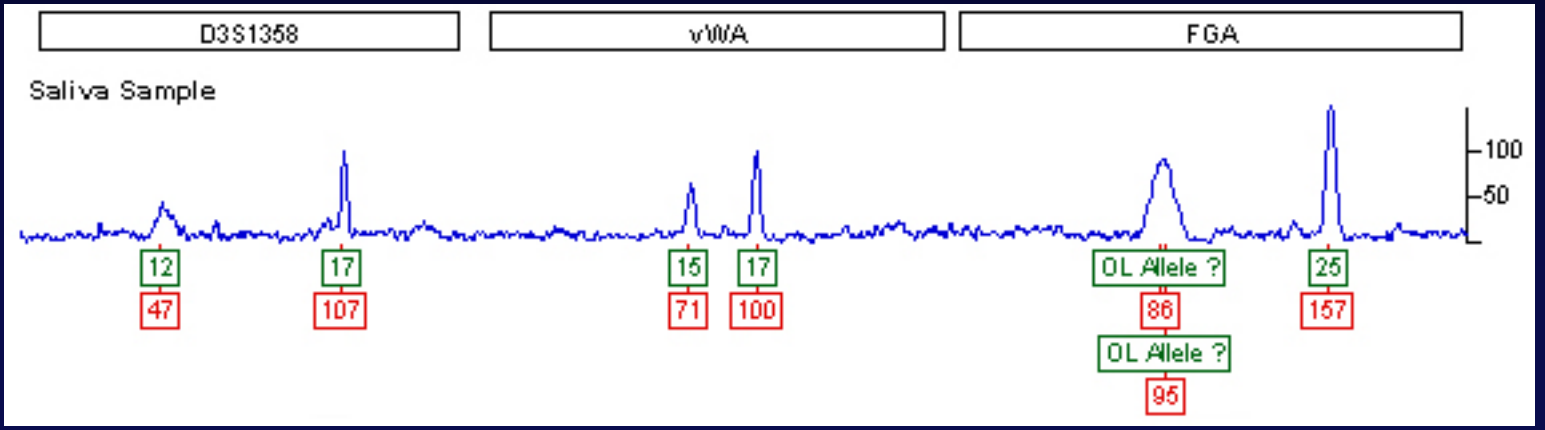


Can “Dick” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25

No -- stochastic effects explain peak height disparity in D3; blob in FGA masks 20 allele.

Opportunities for subjective interpretation?

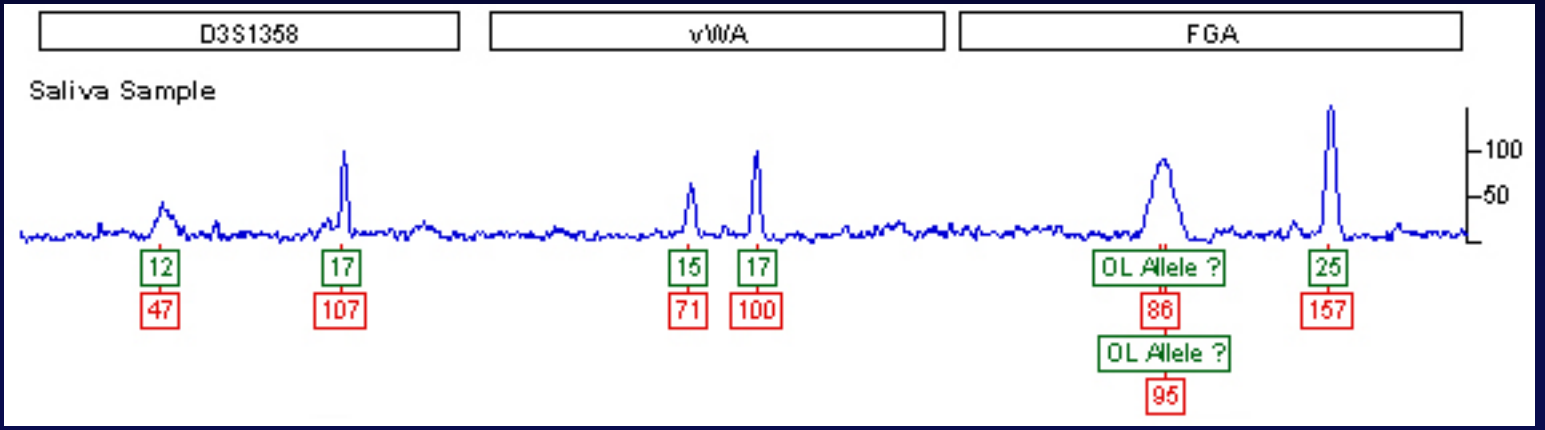


Can “Harry” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25
Harry	14, 17	15, 17	20, 25

No -- the 14 allele at D3 may be missing due to “allelic drop out”; FGA blob masks the 20 allele.

Opportunities for subjective interpretation?



Can “Sally” be excluded?

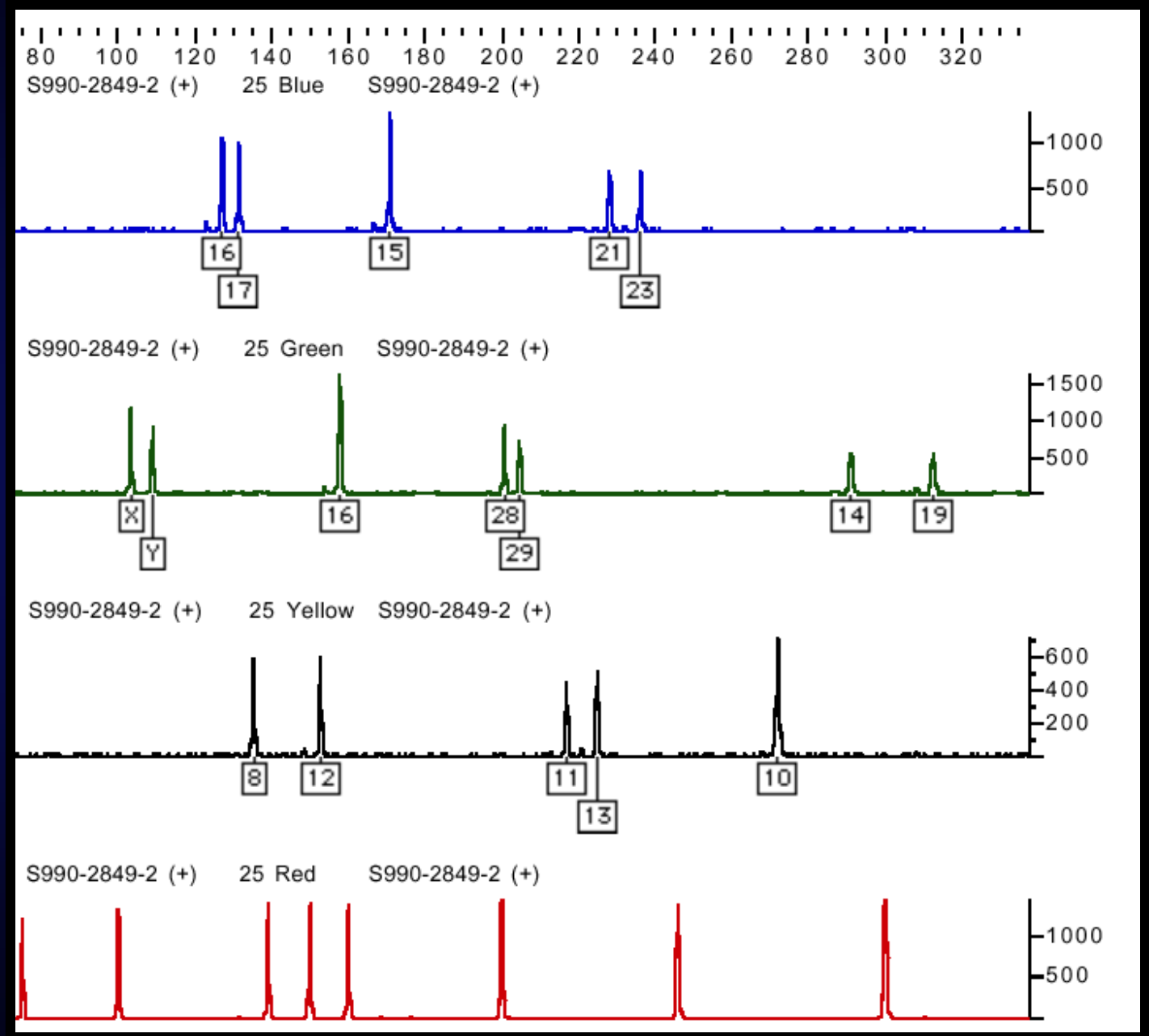
Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25
Harry	14, 17	15, 17	20, 25
Sally	12, 17	15, 15	20, 22

No -- there must be a second contributor; degradation explains the “missing” FGA allele.

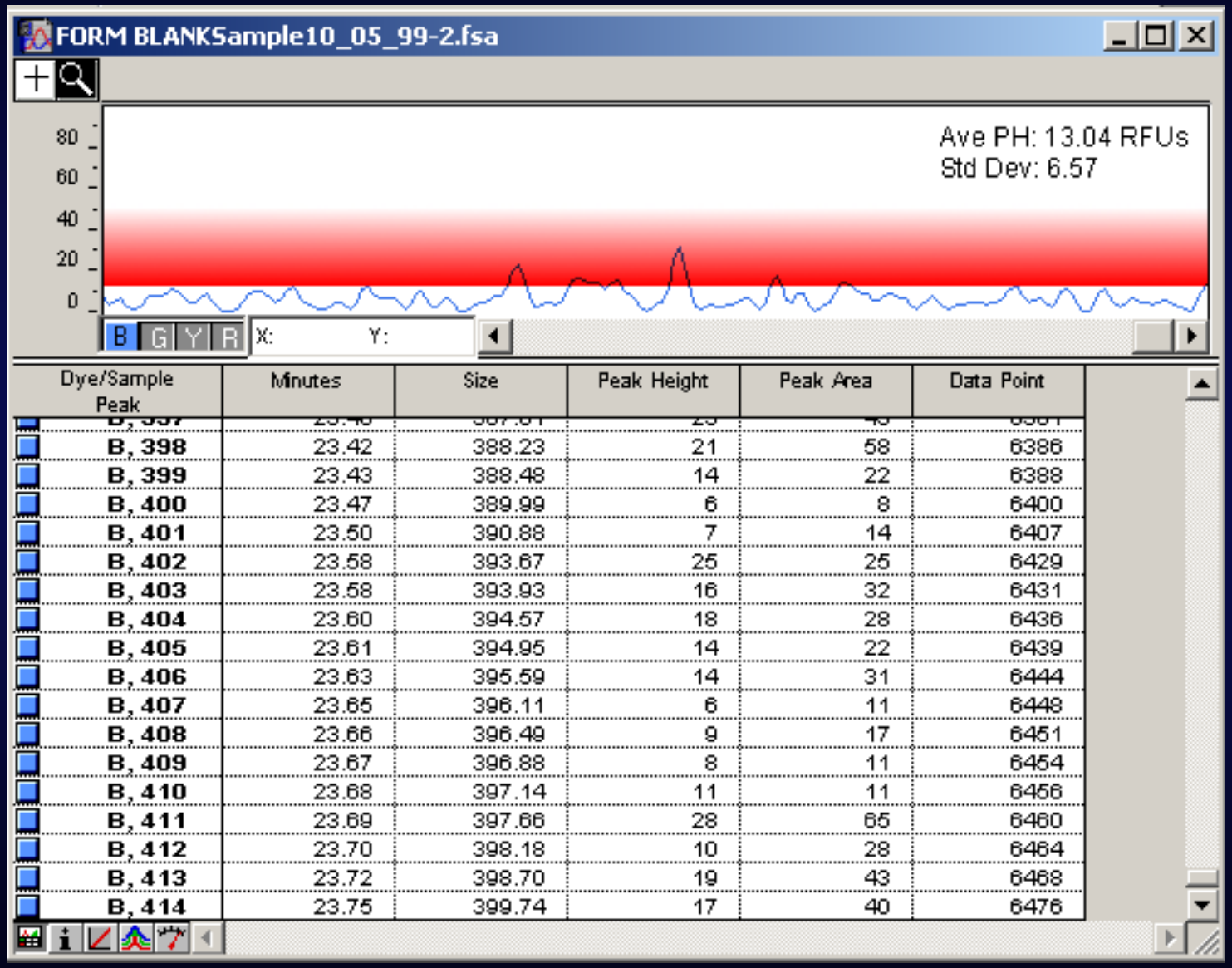
What can be done to make DNA testing more objective?

- Distinguishing between signal and noise
- Deducing the number of contributors to mixtures
- Accounting for relatives
- Be mindful of the potential for human error

Many opportunities to measure baseline

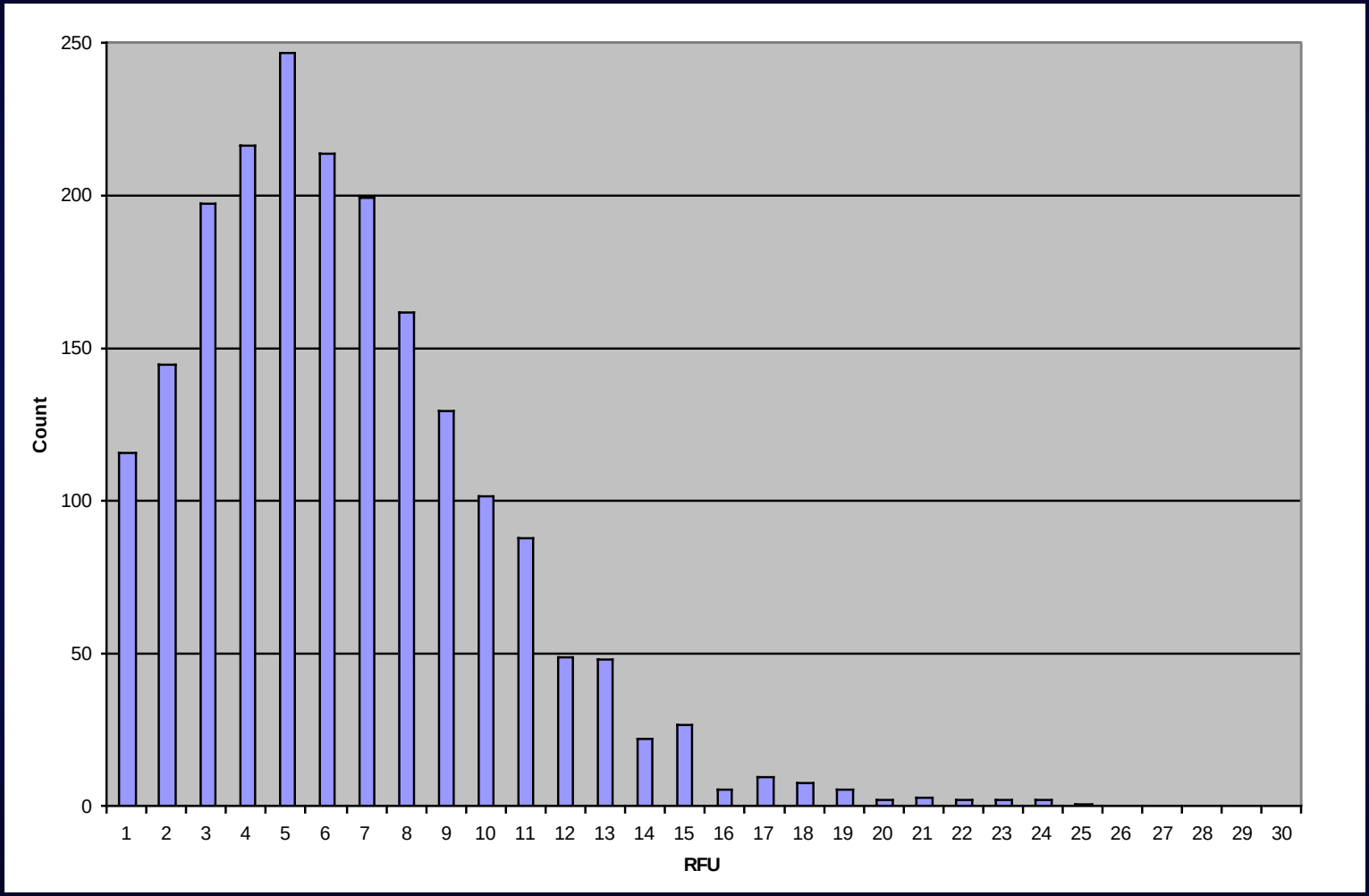


Background noise



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
B, 398	23.42	388.23	21	58	6386
B, 399	23.43	388.48	14	22	6388
B, 400	23.47	389.99	6	8	6400
B, 401	23.50	390.88	7	14	6407
B, 402	23.58	393.67	25	25	6429
B, 403	23.58	393.93	16	32	6431
B, 404	23.60	394.57	18	28	6436
B, 405	23.61	394.95	14	22	6439
B, 406	23.63	395.59	14	31	6444
B, 407	23.65	396.11	6	11	6448
B, 408	23.66	396.49	9	17	6451
B, 409	23.67	396.88	8	11	6454
B, 410	23.68	397.14	11	11	6456
B, 411	23.69	397.66	28	65	6460
B, 412	23.70	398.18	10	28	6464
B, 413	23.72	398.70	19	43	6468
B, 414	23.75	399.74	17	40	6476

RFU levels at all non-masked data collection points

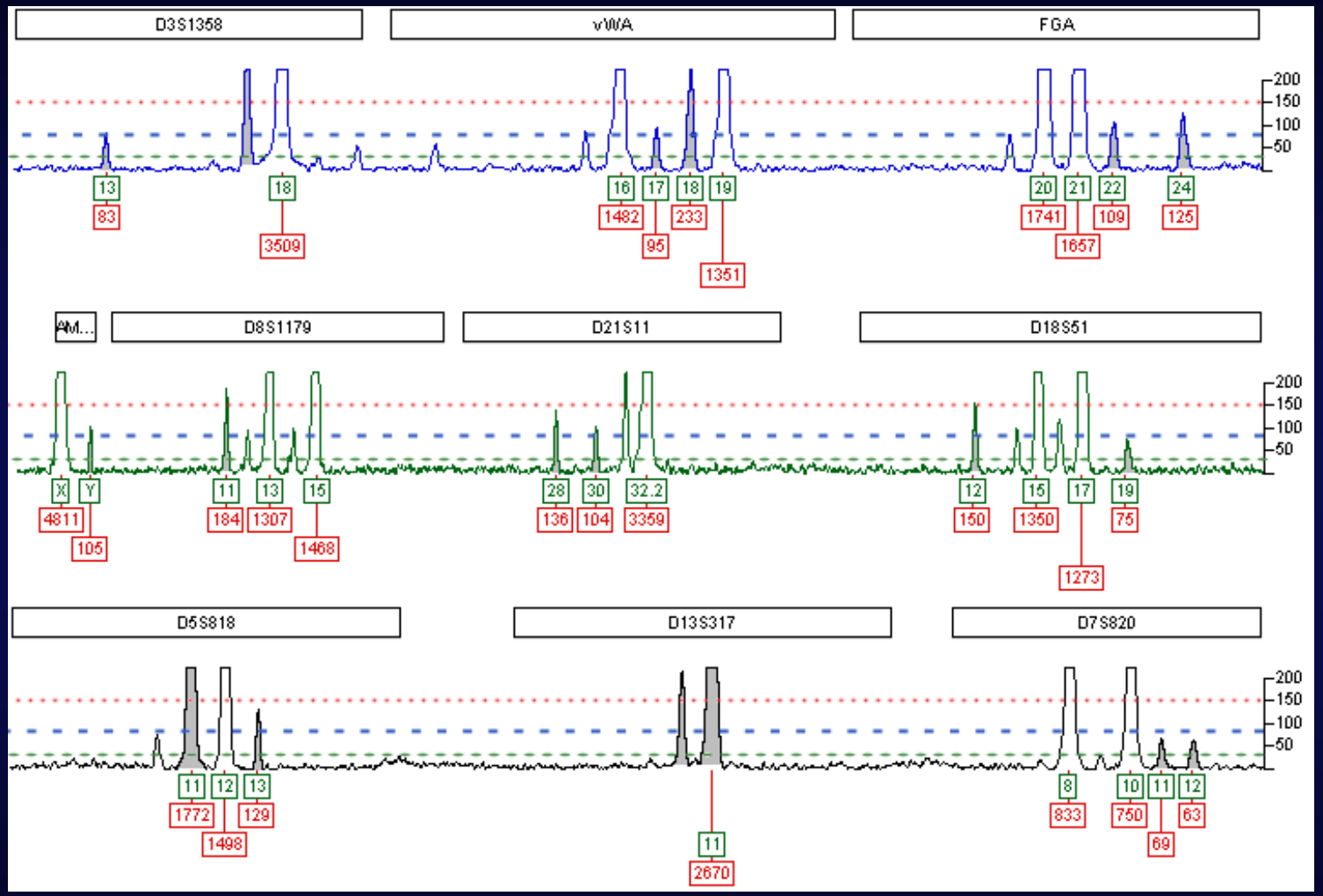


Variation in baseline noise levels

Positive Control		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	6.7	6.9	27.4	75.7
	Average	5.0	3.7	16.1	42.0
	Minimum	3.7	2.4	10.9	27.7
Negative Control		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	13.4	13.2	53.0	145.4
	Average	5.4	3.9	17.1	44.4
	Minimum	4.0	2.6	11.8	30.0
Reagent Blank		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	6.5	11.0	39.5	116.5
	Average	5.3	4.0	17.3	45.3
	Minimum	4.0	2.6	11.8	30.0
All three controls averaged		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	7.1	7.3	29.0	80.1
	Average	5.2	3.9	16.9	44.2
	Minimum	3.9	2.5	11.4	28.9

Average (μ_b) and standard deviation (σ_b) values with corresponding LODs and LOQs from positive, negative and reagent blank controls in 50 different runs. BatchExtract: ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/

Lines in the sand: a two-person mix?

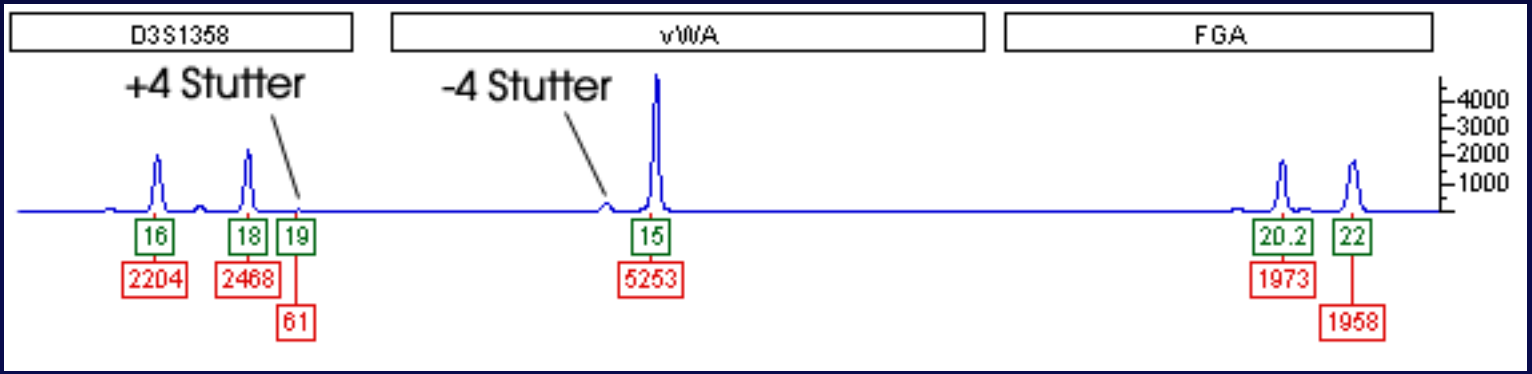


Two reference samples in a 1:10 ratio (male:female). Three different thresholds are shown: 150 RFU (red); LOQ at 77 RFU (blue); and LOD at 29 RFU (green). Gilder et al., January 2007 *JFS*.

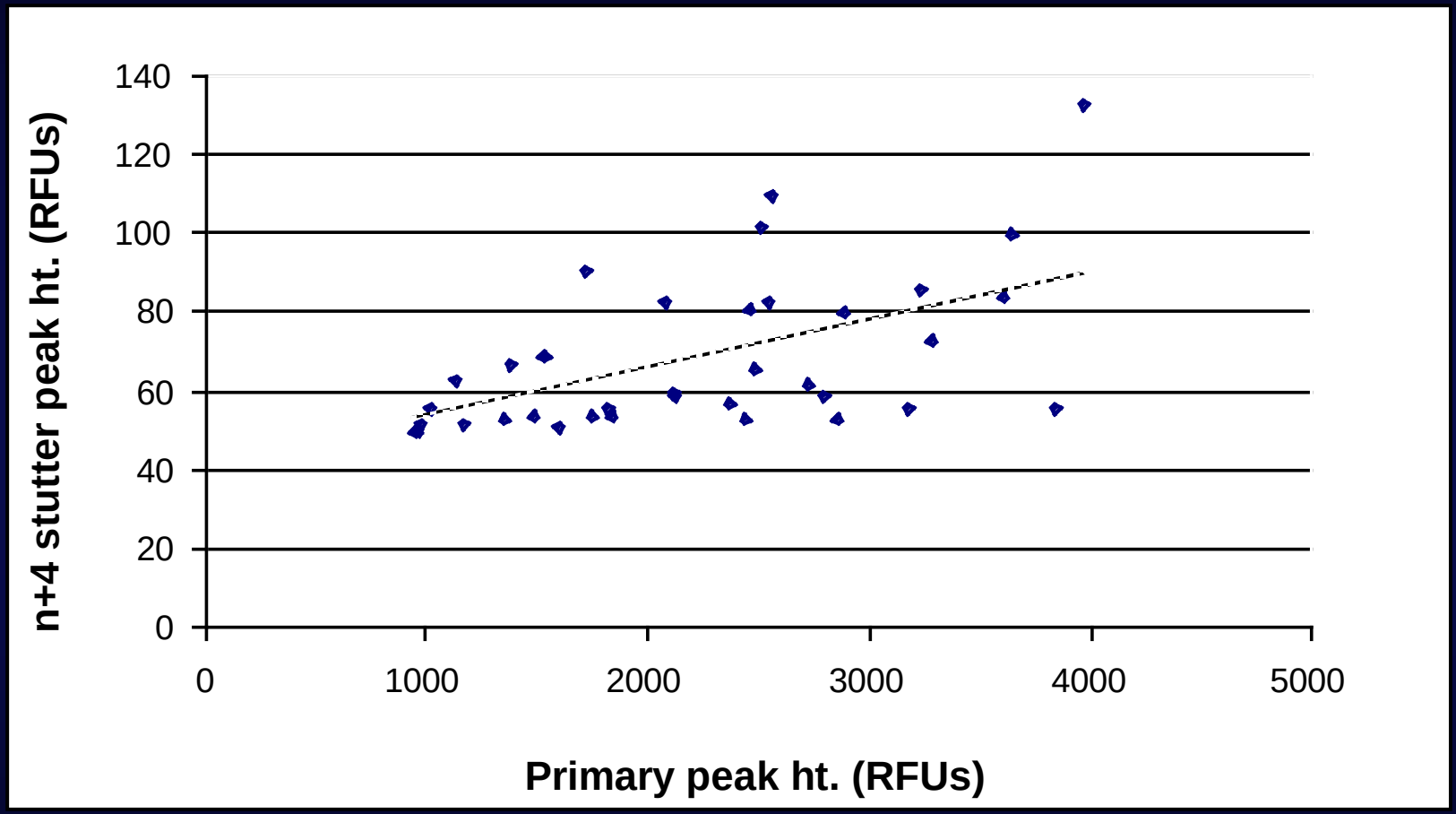
Not all signal comes from DNA associated with an evidence sample

- Stutter peaks
- Pull-up (bleed through)
- Spikes and blobs

Stutter peaks

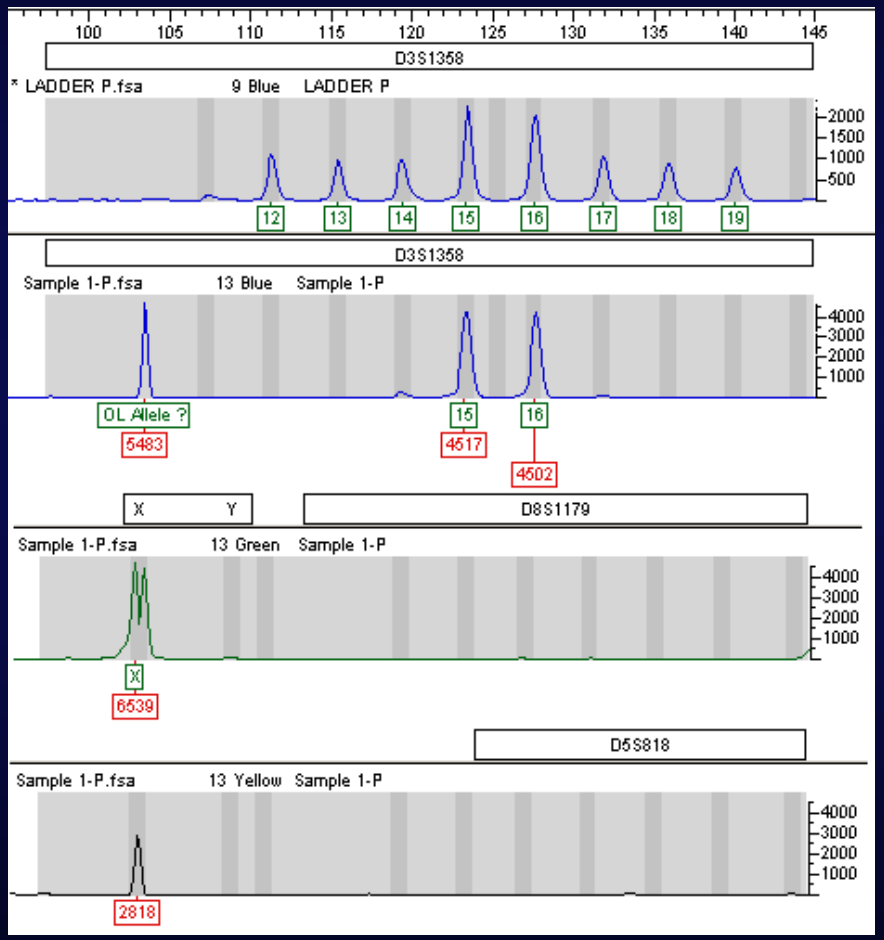


The reality of n+4 stutter

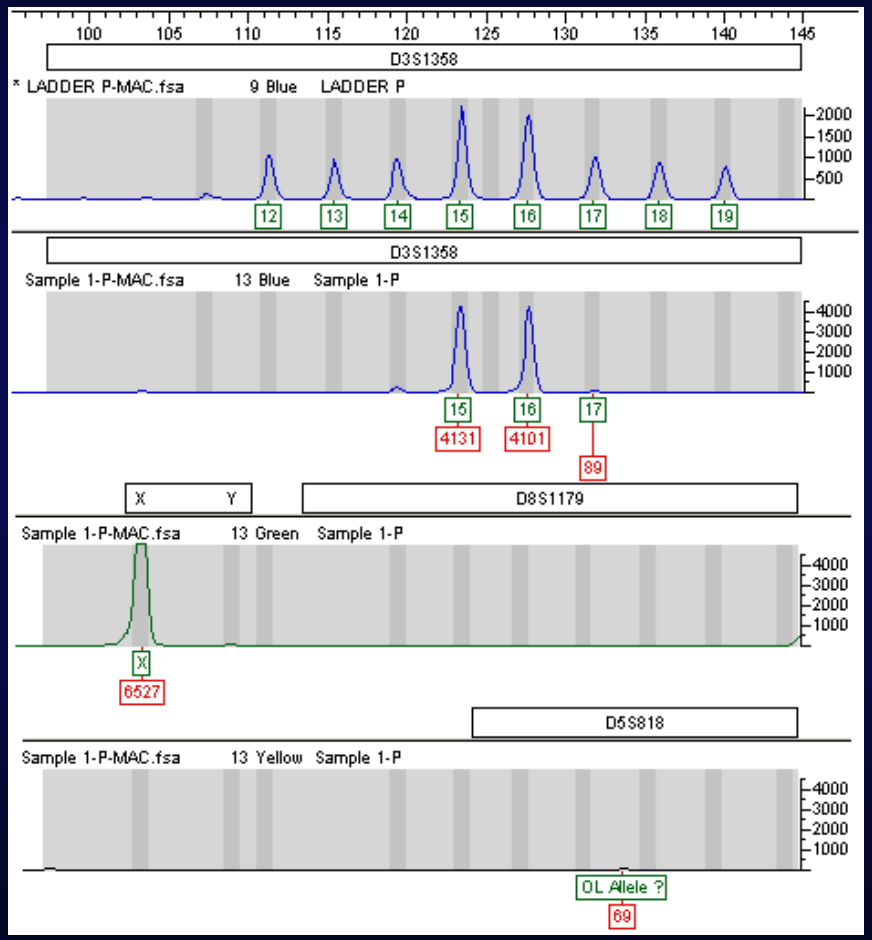


Primary peak height vs. n+4 stutter peak height. Evaluation of 37 data points, R^2=0.293, p=0.0005. From 224 reference samples in 52 different cases. A filter of 5.9% would be conservative. Rowland and Krane, accepted with revision by *JFS*.

Pull-up (and software differences)

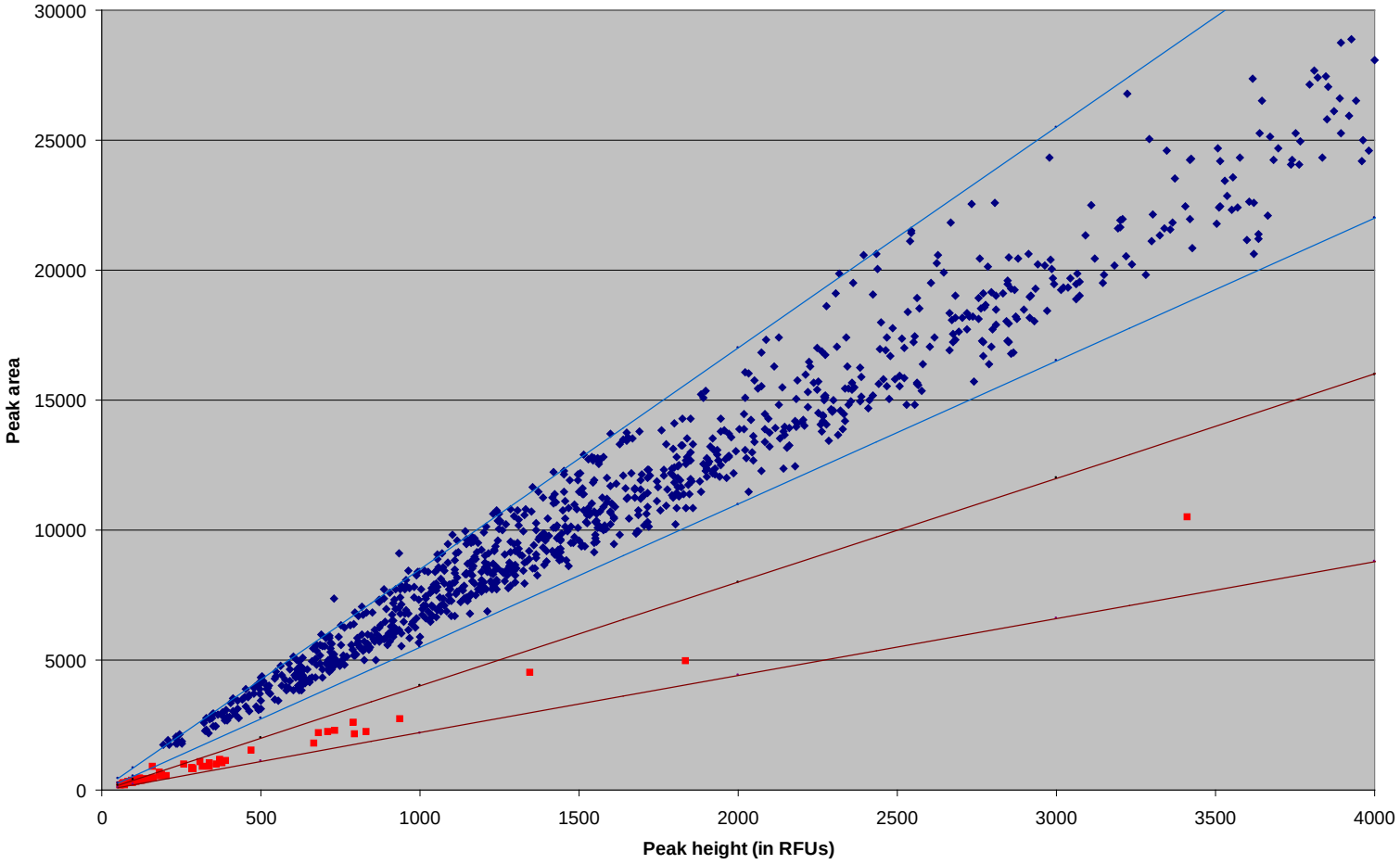


Advanced



Classic

Spikes

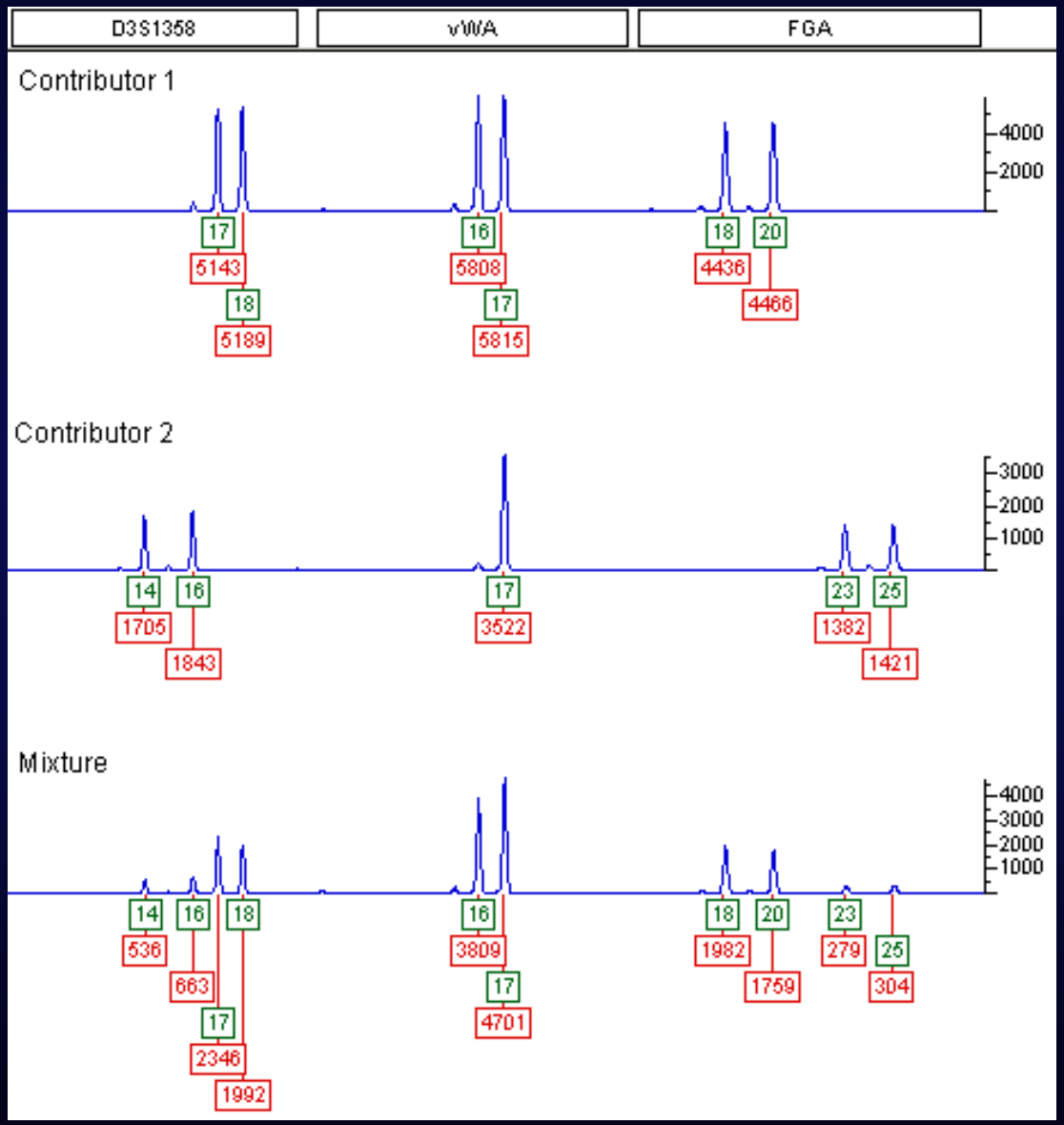


- 89 samples (references, pos controls, neg controls)
- 1010 “good” peaks
- 55 peaks associated with 24 spike events
- 95% boundaries shown

What can be done to make DNA testing more objective?

- Distinguishing between signal and noise
- Deducing the number of contributors to mixtures
- Accounting for relatives
- Be mindful of the potential for human error

Mixed DNA samples



How many contributors to a mixture? If analysts can discard a locus?

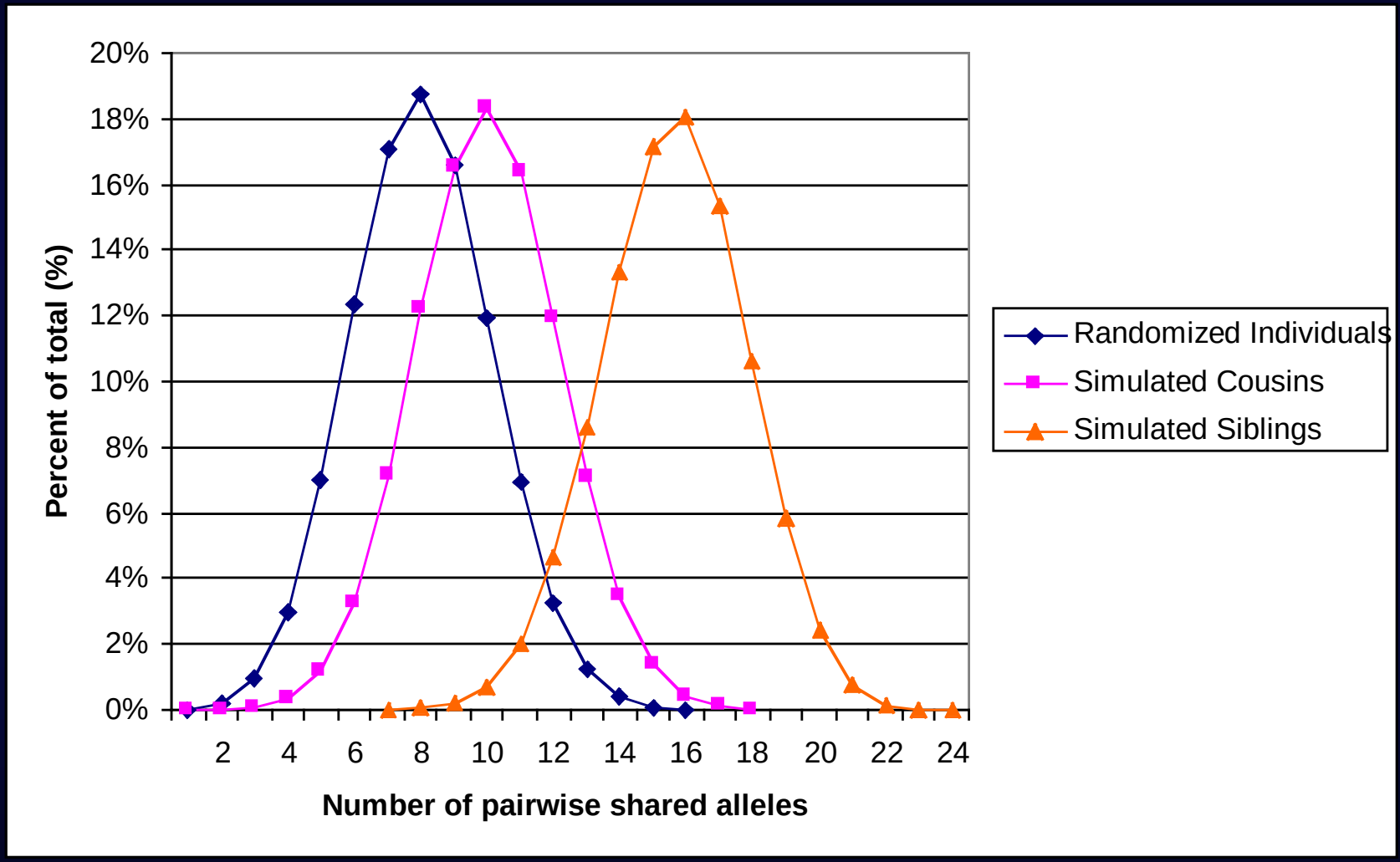
Maximum # of alleles observed in a 3 person mixture	# of occurrences	Percent of cases
2	0	0.00
	8,151	0.02
3	310	0.00
	11,526,219	25.53
4	2,498,139	5.53
	32,078,976	71.07
5	29,938,777	66.32
	1,526,550	3.38
6	12,702,670	28.14

There are 45,139,896 possible different 3-way mixtures of the 648 individuals in the MN BCI database (Paoletti et al., November 2005 *JFS*).

What can be done to make DNA testing more objective?

- Distinguishing between signal and noise
- Deducing the number of contributors to mixtures
- Accounting for relatives
- Be mindful of the potential for human error

Accounting for relatives



Familial searches

- 2003 North Carolina performed post-conviction DNA testing on evidence from a 1984 rape and murder
- Exonerated Darryl Hunt, who had served 18 years of a life sentence
- Database search yielded best match to Anthony Brown with 16/26 alleles
- Brother Willard Brown tested and found to be a perfect match

Thresholds for similarity

- Virginia: “be very, very close”
- California: “appear useful”
- Florida: match at least 21 out of 26 alleles
- North Carolina: 16 out of 26 is enough

Is 16/26 close enough?

- How many pairs of individuals match at 16+ alleles in the previous experiments with unrelated databases of size...

- 1,000: 562 pairs of individuals

- 5,000: 13,872 pairs of individuals

- 10,000: 52,982 pairs of individuals

Is the true DNA match a sibling or a random individual?

- Given a closely matching profile, who is more likely to match, a sibling or a randomly chosen, unrelated individual?

- Use a likelihood ratio (Paoletti et al., Winter 2006 *Jurimetrics*)

$$LR = \frac{P(E \mid sib)}{P(E \mid random)}$$

Probabilities of siblings matching at 0, 1 or 2 alleles

- Weir and NRC I only present probabilities that siblings match perfectly.

$$P(E \mid sib) = \begin{cases} \dfrac{P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 0 \\ \dfrac{P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 1 \\ \dfrac{1 + P_a + P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 2 \end{cases}$$

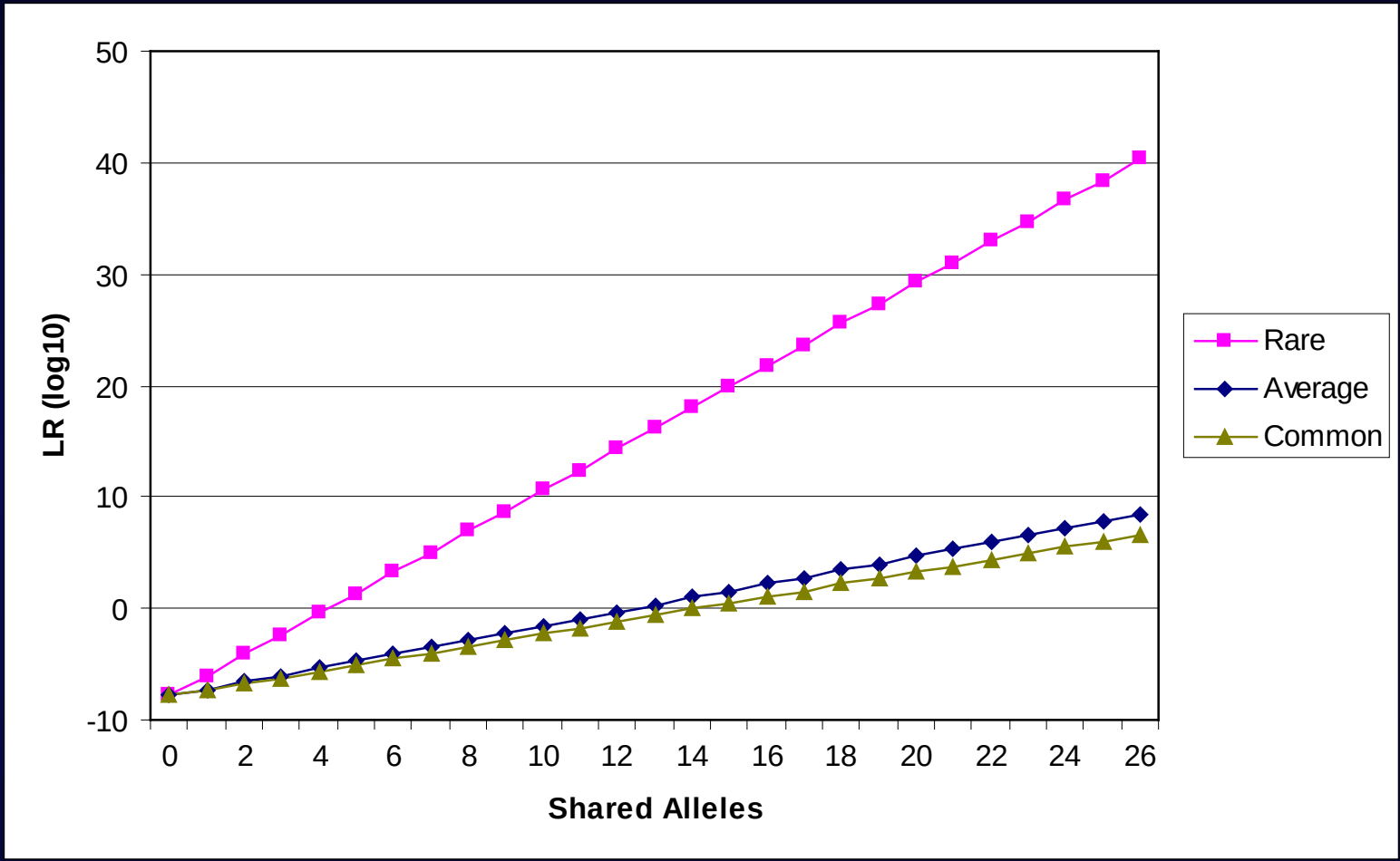
1

Probabilities of parent/child matching at 0, 1 or 2 alleles

- Weir and NRC I only present probabilities that parent/child match perfectly.

$$P(E \mid parent/child) = \begin{cases} 0, & if \quad shared = 0 \\ \frac{P_b}{2}, & if \quad shared = 1 \\ \frac{P_a + P_b}{2}, & if \quad shared = 2 \end{cases}$$

Considering rarity of alleles



- As few as 5/26 rare alleles
- 13/26 average alleles
- 15/26 common alleles

Thresholds for similarity

- Virginia: “be very, very close”
- California: “appear useful”
- Florida: match at least 21 out of 26 alleles
- North Carolina: 16 out of 26 is enough

CODIS search simulation

Relationship	Average alleles	Std dev alleles	Average loci	Std dev loci	CODIS High	CODIS Medium	CODIS Low	20+ allele matches
Siblings	16.6	2.3	11.6	1.1	3.21E-06	9	2349	946
Parent-Child	15.8	1.5	13.0	0.0	1.46E-09	1	10000	96
Half-sibling	12.3	2.1	10.3	1.4	9.37E-12	0	466	1
Cousins	10.5	2.2	8.9	1.6	2.72E-13	0	70	0
Uncle/Nephew	12.3	2.1	10.3	1.4	9.40E-12	0	464	3
Grandparent-Grandchild	12.3	2.1	10.3	1.5	9.37E-12	0	496	3
Unrelated	8.7	2.2	7.6	1.7	1.61E-15	0	10	0

10,000 pairs of each group

Likelihood ratio approach

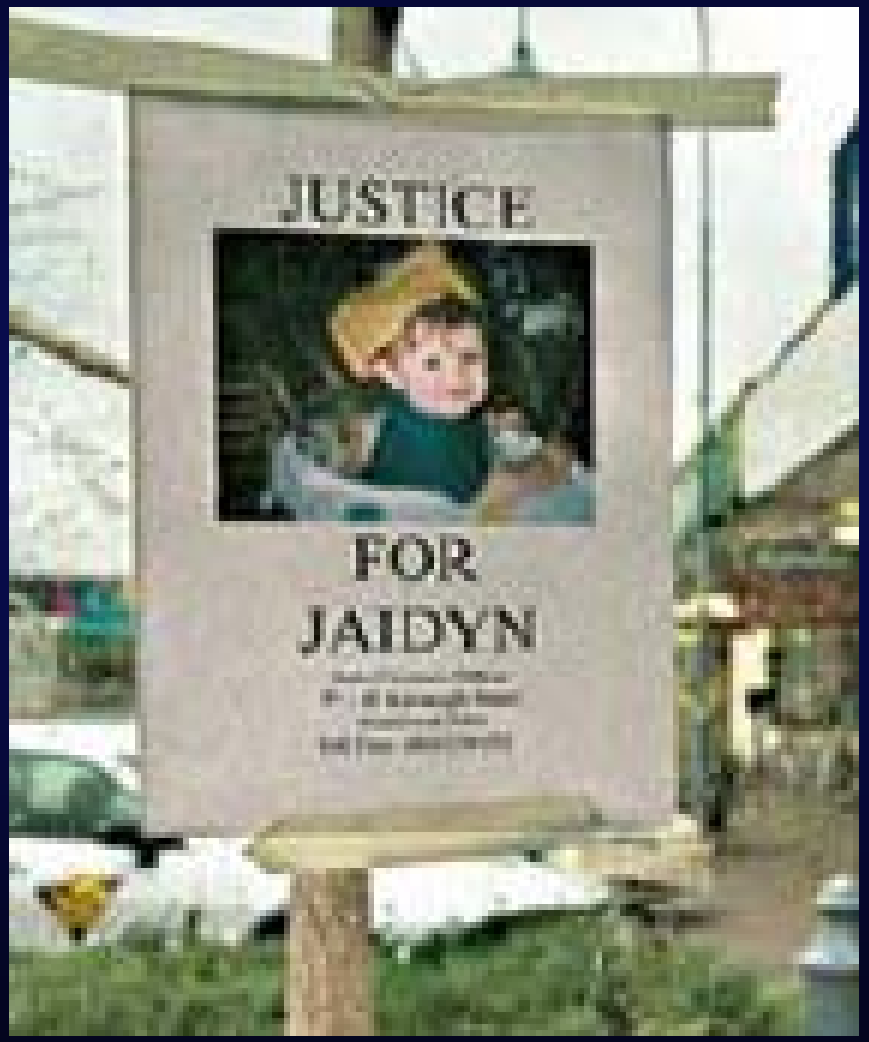
Relationship	LR > 1	LR > 10000
Actual Siblings : Unrelated	9967	4590
Actual Parent/Child : Unrelated	9999	2807
Actual Half-Siblings : Unrelated	7566	1
Actual Cousins : Unrelated	5723	0
Actual Uncle/Nephew : Unrelated	7565	5
Actual Grandparent/Grandchild : Unrelated	7562	2
Incorrectly assumed siblings : Actual unrelated	201	0
Incorrectly assumed parent/child : Actual unrelated	10	0
Incorrectly assumed uncle/half-sib/grandparent : Actual unrelated	1096	0
Incorrectly assumed cousin : Actual unrelated	2125	0
Incorrectly assumed sibling : Actual parent/child	622	0
Incorrectly assumed parent/child : Actual sibling	1000	0

What can be done to make DNA testing more objective?

- Distinguishing between signal and noise
- Deducing the number of contributors to mixtures
- Accounting for relatives
- Be mindful of the potential for human error

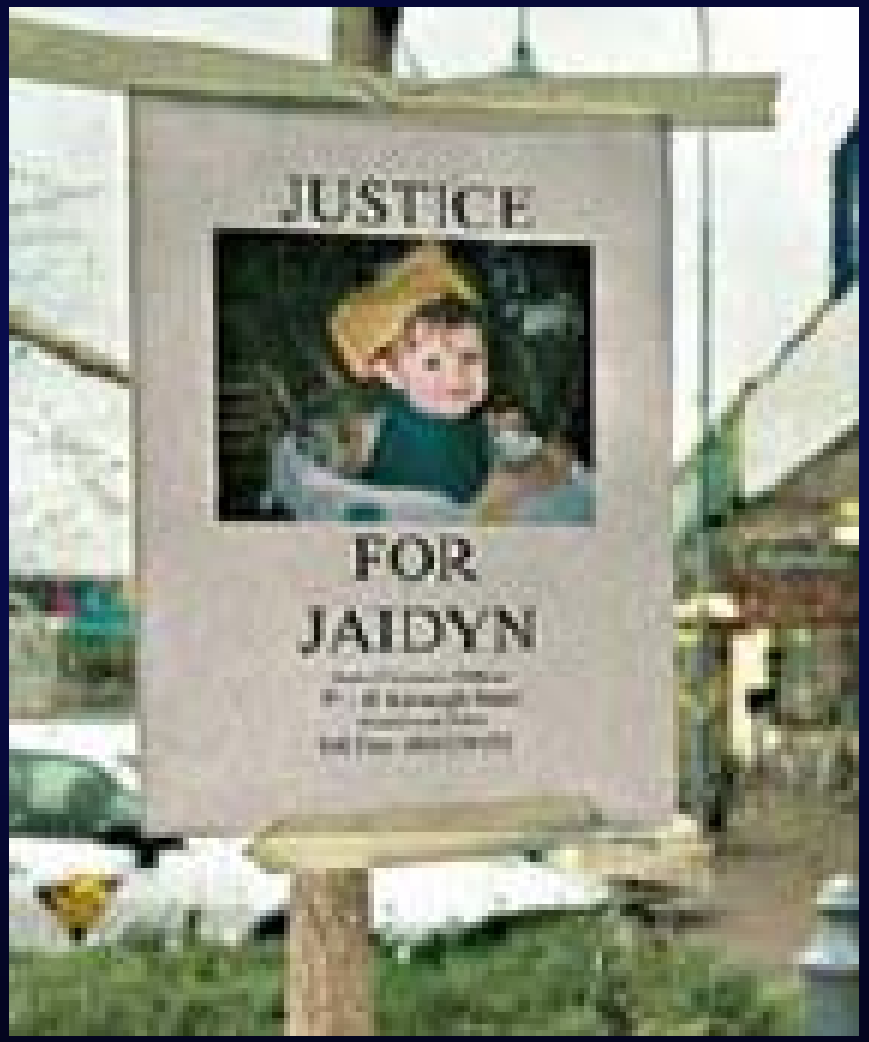
Victorian Coroner's inquest into the death of Jaidyn Leskie

- Toddler disappears in bizarre circumstances: found dead six months later
- Mother's boy friend is tried and acquitted.
- Unknown female profile on clothing.
- Cold hit to a rape victim.
- RMP: 1 in 227 million.
- Lab claims "adventitious match."

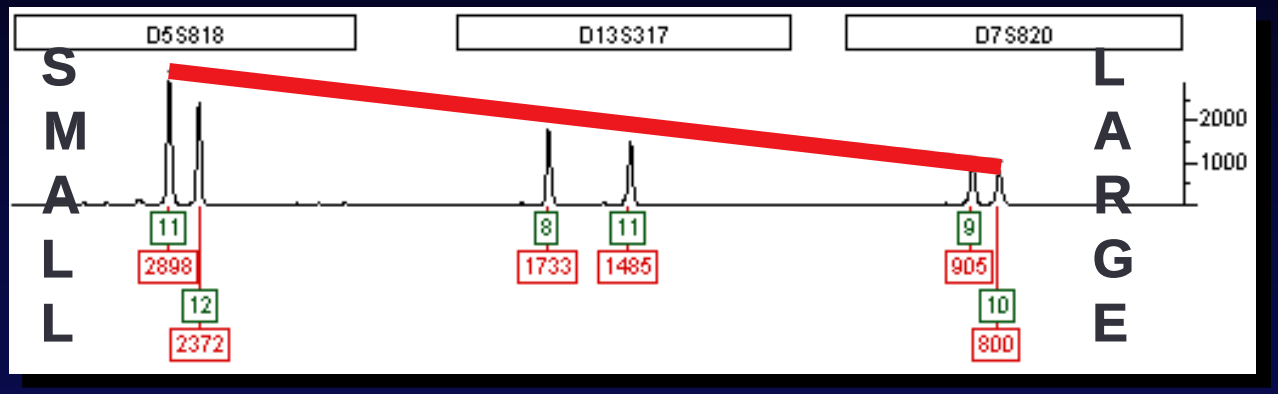


Victorian Coroner's inquest into the death of Jaidyn Leskie

- Condom with rape victim's DNA was processed in the same lab 1 or 2 days prior to Leskie samples.
- Additional tests find matches at 5 to 7 more loci.
- Review of electronic data reveals low level contributions at even more loci.
- Degradation study further suggests contamination.



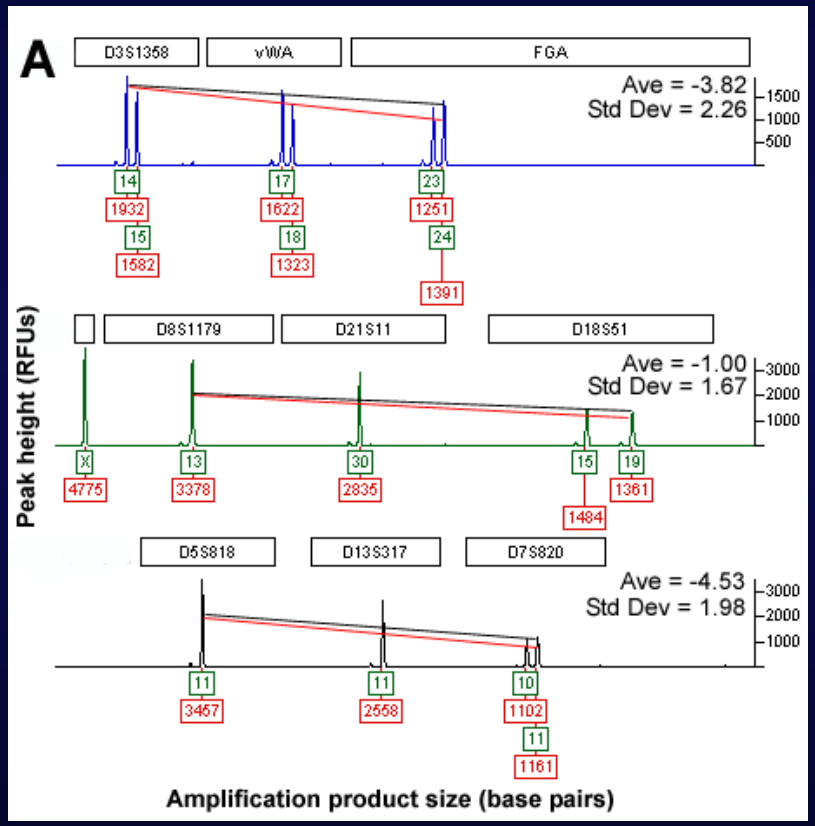
Degradation, inhibition



- When biological samples are exposed to adverse environmental conditions, they can become degraded
  - **Warm, moist, sunlight, time**
- Degradation breaks the DNA at random
- Larger amplified regions are affected first
- Classic ‘ski-slope’ electropherogram
- Degradation and inhibition are unusual and noteworthy.

Degradation, inhibition

The Leskie Inquest, a practical application

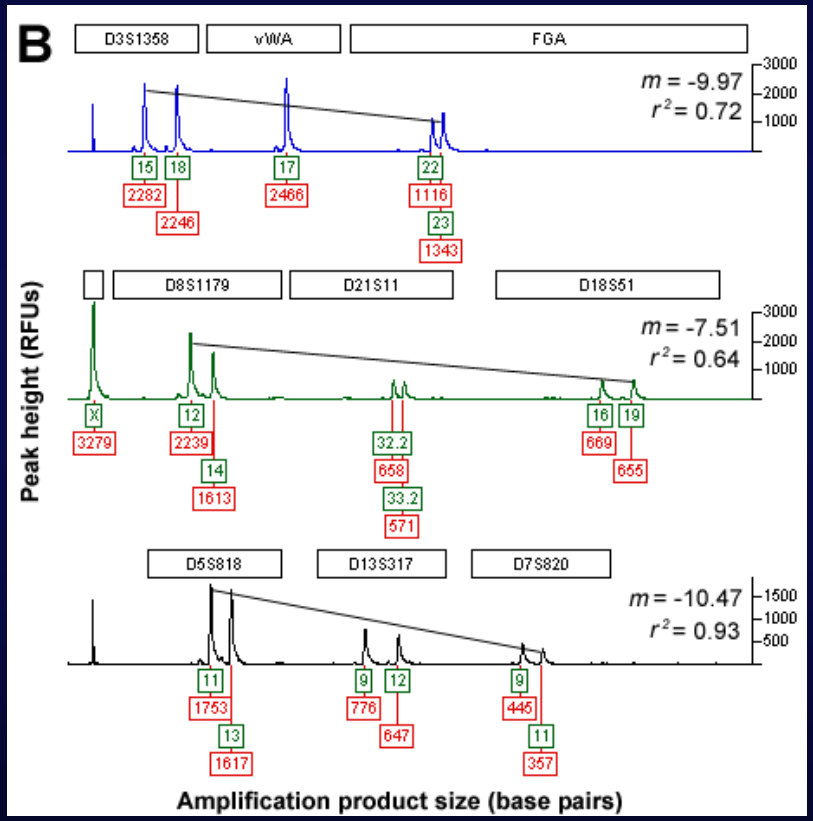


- Undegraded samples can have “ski-slopes” too.
- How negative does a slope have to be to an indication of degradation?
- Experience, training and expertise.
- Positive controls should not be degraded.

Degradation, inhibition

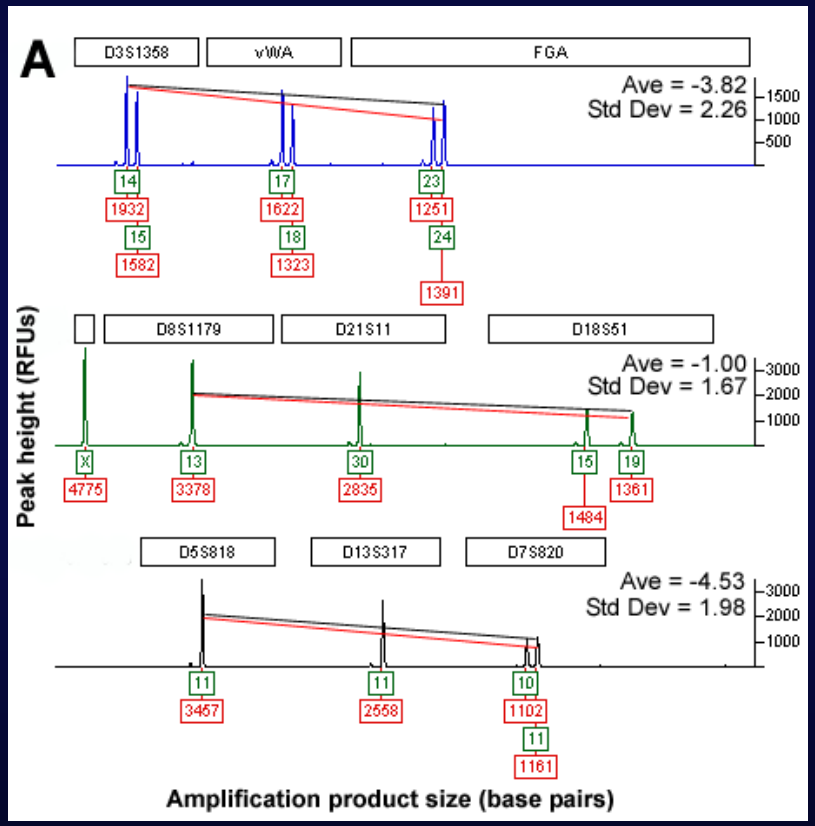
The Leskie Inquest

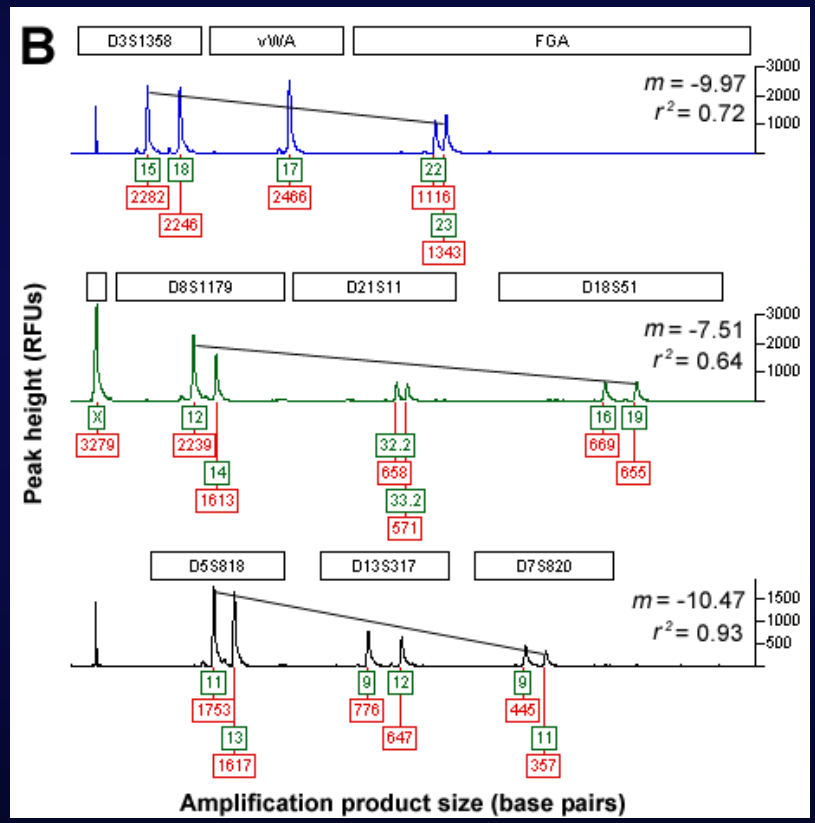
- DNA profiles in a rape and a murder investigation match.
- Everyone agrees that the murder samples are degraded.
- If the rape sample is degraded, it could have contaminated the murder samples.
- Is the rape sample degraded?



Degradation, inhibition

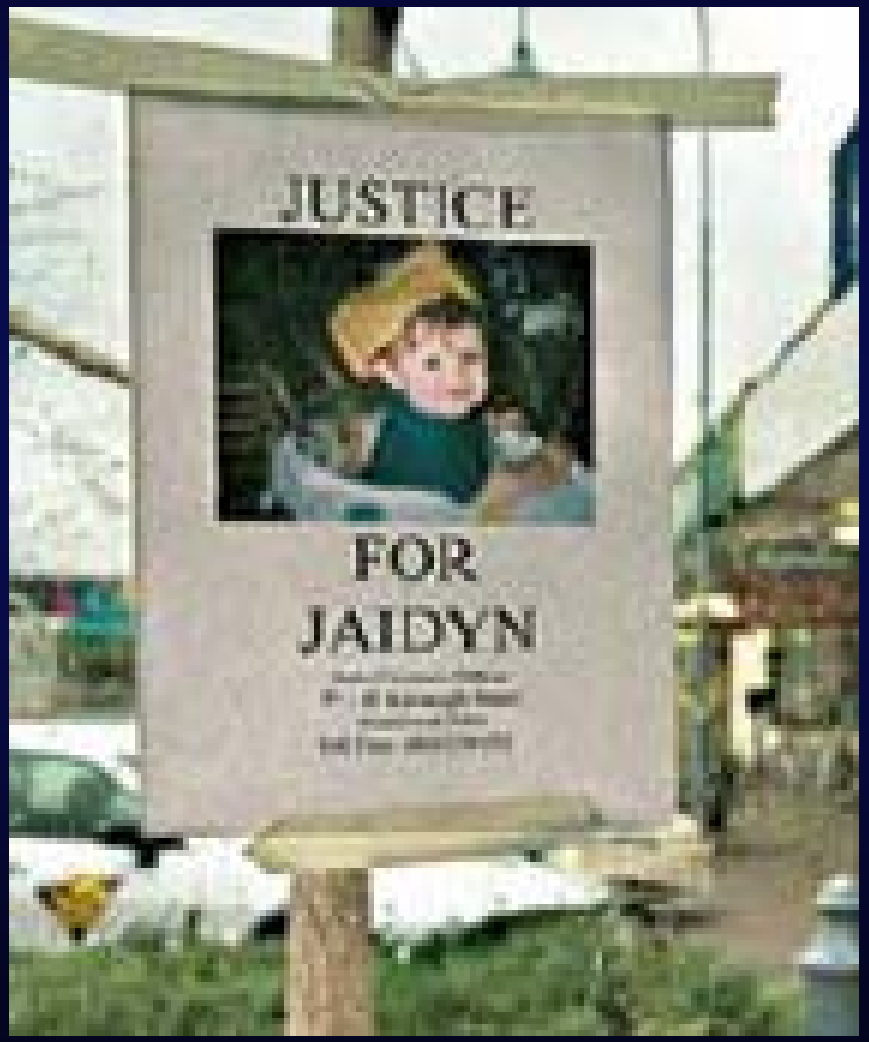
The Leskie Inquest





Victorian Coroner's inquest into the death of Jaidyn Leskie

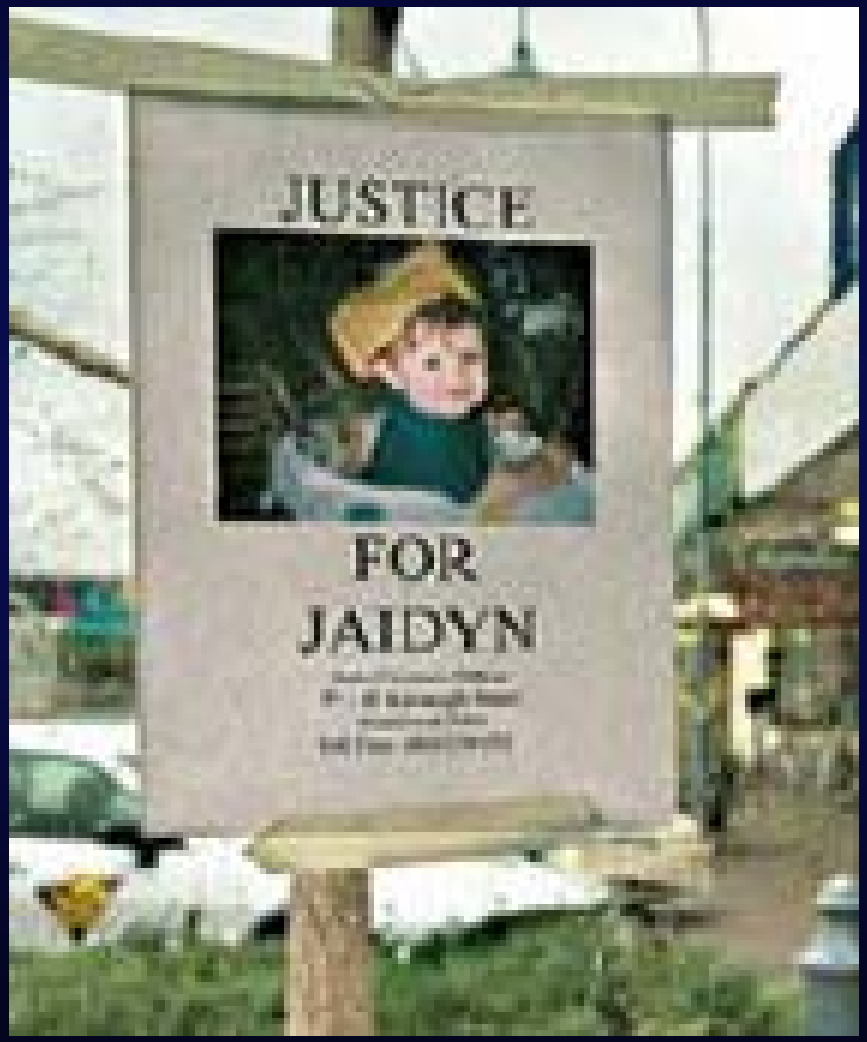
"8. During the conduct of the preliminary investigation (before it was decided to undertake an inquest) the female DNA allegedly taken from the bib that was discovered with the body was matched with a DNA profile in the Victorian Police Forensic Science database. This profile was from a rape victim who was subsequently found to be unrelated to the Leskie case."



Victorian Coroner's inquest into the death of Jaidyn Leskie

"8. The match to the bib occurred as a result of contamination in the laboratory and was not an adventitious match. The samples from the two cases were examined by the same scientist within a close time frame."

www.bioforensics.com/articles/Leskie_decision.pdf



The science of DNA profiling is sound.

But, not all of DNA profiling is science.

This is especially true in situations involving: small amounts of starting material, mixtures, relatives, and analyst judgment calls.

Resources

- Internet
  - **Forensic Bioinformatics Website: http://www.bioforensics.com/**
  - **Applied Biosystems Website: http://www.appliedbiosystems.com/ (see human identity and forensics)**
  - **STR base: http://www.cstl.nist.gov/biotech/strbase/ (very useful)**
- Books
  - **'Forensic DNA Typing' by John M. Butler (Academic Press)**
- Scientists
  - **Larry Mueller (UC Irvine)**
  - **Simon Ford (Lexigen, Inc. San Francisco, CA)**
  - **William Shields (SUNY, Syracuse, NY)**
  - **Mike Raymer and Travis Doom (Wright State, Dayton, OH) Marc Taylor (Technical Associates, Ventura, CA)**
  - **Keith Inman (Forensic Analytical, Haywood, CA)**
- Testing laboratories
  - **Technical Associates (Ventura, CA)**
  - **Forensic Analytical (Haywood, CA)**
- Other resources
  - **Forensic Bioinformatics (Dayton, OH)**